Power and Sample size Estimation for Bioequivalence Studies

Short cursory excerpt

D. Labes (July 2010)

The used matematical and statistical apparatus is here only formulated for the evaluation of the pharmacokinetic metrics which are assumed log-normal distributed.
The description follows closely Diletti, E., Hauschke, D., and Steinijans, V.W. (1991).

For the formulas using untransformed PK metrics refer to Phillips (1990).

The TOST procedure

Let μ_T and μ the expected mean values of the pharmacokinetic metric (f.i. AUC, or Cmax) of the Test and Reference formulation to be compared within a bioequivalence study.
Let the interval (Θ_1, Θ_2) denote the bioequivalence acceptance range where $0<\Theta_1<1<\Theta_2$.
Most regulatory guidances set $(\Theta_1, \Theta_2) = (0.8, 1.25)$ for log-normal distributed pharmaco-kinetic metrics (AUC, Cmax). Other values may be used (f.i. (0.75, 1.3333) for widened Cmax, or (0.9, 1.1111) for NTI drugs).

The bioequivalence test problem then becomes:

$$H_0 : \mu_T / \mu_R \leq \Theta_1 \text{ or } \mu_T / \mu_R \geq \Theta_2 \qquad \text{(bioinequivalence)}$$

$$H_1 : \Theta_1 < \mu_T / \mu_R < \Theta_2 \qquad \text{(bioequivalence)}$$

In case of log-normal distributed pharmaco-kinetic metrics the test problem is transformed accordingly to

$$H_0 : \log(\mu_T / \mu_R) \leq \log(\Theta_1) \text{ or } \log(\mu_T / \mu_R) \geq \log(\Theta_2) \qquad \text{(bioinequivalence)}$$

$$H_1 : \log(\Theta_1) < \log(\mu_T / \mu_R) < \log(\Theta_2) \qquad \text{(bioequivalence)}$$

were log() denotes the natural logarithmn. $\log(\mu_T / \mu_R) = \log(\mu_T) - \log(\mu_R)$ is estimated by the difference of the arithmetic means of the log-transformed observations $\overline{X}_T - \overline{X}_R$.

H_0 is rejected in favor of bioequivalence if the classical $(1-2\alpha)100\%$ confidence interval for μ_T/μ_R is included in the bioequivalence range (Westlake 1981, 1988).
The inclusion of the $(1-2\alpha)100\%$ confidence interval in the acceptance range is equivalent to the two one-sided t-test (Schuirmann 1987).

Bioequivalence is in case of data from a 2x2 cross-over study concluded if the following two conditions hold true:

$$t_1 = \frac{\overline{X}_T - \overline{X}_R - \log(\Theta_1)}{s_e\sqrt{2/n}} \geq t(1-\alpha, n-2)$$

$$t_2 = \frac{\overline{X}_T - \overline{X}_R - \log(\Theta_2)}{s_e\sqrt{2/n}} \leq t(1-\alpha, n-2)$$

where s_e is estimated from the mean squared error of an appropriate analysis of variance. $t(1-\alpha, n-2)$ denotes the (1- α)-quantile of the central t-distribution with n-2 degrees of freedom. n is the number of subjects under study.

Here we assume that the number of subjects within the two sequence groups TR and RT, respectively, are the same. This assumption is also applied for the other designs covered within the package *PowerTOST*.

Power of the TOST procedure

The power of a statistical test is the probability that the hypothesis H_0, in our case bioineqivalence, is rejected if the alternative hypothesis H_1, here bioequivalence, is true. In other words the probability of correctly accepting bioequivalence is the power of the test. The power of the two one-sided t-tests (TOST) is thus given by

$$Power = \mathrm{Pr}\,ob\bigl(t_1 \geq t_{(1-\alpha, n-2)} \ and \ t_2 \leq -t_{(1-\alpha, n-2)} \,|\, bioequivalence \ holds\bigr) \qquad \textbf{(I)}$$

The t_1 and t_2 values are the t-test statistics of the two one-sided t-tests descibed above.

Owen (1965) showed that (t_1, t_2) has a bivariate non-central *t*-distribution and that the power can be calculated as the difference of two definite integrals (Owen's Q function):

$$Power = 1 - \beta = Q_{df}(-t_{(1-\alpha, n-2)}, \delta_2; 0, R) - Q_{df}(t_{(1-\alpha, n-2)}, \delta_1; 0, R) \qquad \textbf{(II)}$$

where $t_{(1-\alpha,\,df)}$ is the (1-α) quantile of a *t*-distribution with df degrees of freedom. df is (n-2) in case of a classical 2x2 cross-over design and

$$\Theta_0 = null\ ('true')\ ratio \qquad \textbf{(IIa)}$$

$$\delta_1 = \frac{\log(\Theta_0) - \log(\Theta_1)}{s_e\sqrt{2/n}}$$

$$\delta_2 = \frac{\log(\Theta_0) - \log(\Theta_2)}{s_e\sqrt{2/n}}$$

$$R = \frac{\sqrt{df} \cdot (\delta_1 - \delta_2)}{2 \cdot t_{(1-\alpha, df)}}$$

for log-transformed pharmaco-kinetic metrics, where s_e is the residual standard error, Θ_1 and Θ_2 are the lower and upper bioeqivalence acceptance bounds (usually 0.8 and 1.25).

The residual variance (s^2_e) is connected to the within-subject coefficient of variation CV by

$$s_e^2 = mse = \log(CV^2 + 1)$$

$$CV = \sqrt{\exp(s_e^2) - 1}$$

Owen's Q function is defined as:

$$Q_v(t, \delta; a, b) = \frac{\sqrt{2\pi}}{\Gamma\left(\frac{v}{2}\right) \cdot 2^{\frac{(v-2)}{2}}} \int_a^b \Phi\left(\frac{t \cdot x}{\sqrt{v}} - \delta\right) \cdot x^{v-1} \cdot \varphi(x) \cdot dx \qquad \textbf{(III)}$$

where $\Gamma(x)$ is the gamma-function, $\varphi(x)$ and $\Phi(X)$ are the density and cumulative distribution function of the standard normal distribution, respectively.

Owen's Q function was long part of the SAS system (SAS® Analyst 1999), but undocumented until SAS9.2. It was implemented here in the R package *PowerTOST* via numerical evaluation of the definite integral using the `integrate()` function of the package *stats*, part of the base R distribution (see implementation details below).

Equation (**II**) can be approximated by the univariate non-central t-distribution via

$$Power = pt(-t_{(1-\alpha, n-2)}, n-2, \delta_2) - pt(t_{(1-\alpha, n-2)}, n-2, \delta_1) \qquad \textbf{(IV)}$$

where $pt(t, df, \delta)$ is the distribution function of the non-central t distribution with df degrees of freedom and noncentrality parameter δ.
Equation (III) can further approximated, if the non-central t-distribution is approximated by a "shifted" central t-distribution, according to

$$Power = pt(-\delta_2 - t_{(1-\alpha, n-2)}, n-2) - pt(t_{(1-\alpha, n-2)} - \delta_1, n-2) \qquad \textbf{(V)}$$

where $pt(t, df)$ is the distribution function of the central t-distribution with df degrees of freedom.

Both approximations perform very well if the degrees of freedom v are reasonable high and the obtained power is in the usually interesting range (≥ 60-70%).

Equation (**III**) is used throughout the book from S.A. Joulious (2010), without indicating that it is an aprroximation; and in many other papers.

Equation (**IV**) is used in the book by Chow and Liu (2009) in chapter 9 concerning sample size calculations for higher-order (replicate) crossover designs, also without indicating the approximate nature. It is implemented in the commercial sample size software PASS 2008 (Hintze J. 2008), module "Equivalence of means/Two means in a higher order cross-over design".

Other study designs

The formulas for other study designs used in bioequivalence studies differ from the given ones only by

- the degrees of freedom df and
- the factor 2 under the square root in the denominator of the "non-centrality" parameters δ_1 and δ_2.

The factor 2 has to be replaced by the so-called design constant b_k.

This holds if the same assumptions are made as in the 2x2 cross-over, namely the number of subjects in the sequence groups or the two groups in the parallel group design are equal, the within-subject variabilities or the variabilities in the two parallel groups of the Test and Reference formulations are assumed equal and no subject by formulation interaction is incorporated in the ANOVA for replicate cross-over designs.

See the function `known.designs()` for the values of df=degrees of freedom and b_k implemented.

For the two-group parallel design the sample size is the number of subjects in each treatment group. The CV to be used here is the CV of the total variability.

For the cross-over designs n is the total number of subjects and the CV to be used here is the within-subject CV (CV of the residual error).

Sample size estimation

Equation **(II),** or the approximations (IV) and (V), respectively, are implicit in n – the sample size – and can be solved for given n, alpha, power to achieve, bioequivalence margins and the assumed null ('true') ratio.

The algorithmn starts with a suitable choosen value of the sample size, calculates the power for that and increases / decreases this start value in steps of the sequence groups in the study design until the power reaches or exceed the desired level.

The start value is choosen via the large sample size approximation of the power equation (Joulious 2010), namely the maximum of n_{01} and n_{02} according to

$$n_{01} = 0.5 b_k \left(\frac{s_e \sqrt{2} \left(z_{(1-\alpha)} + z_{(1-\beta)} \right)}{\left(\log(\Theta_0) - \log(\Theta_1) \right)} \right)^2$$

$$n_{02} = 0.5 b_k \left(\frac{s_e \sqrt{2} \left(z_{(1-\alpha)} + z_{(1-\beta)} \right)}{\left(\log(\Theta_0) - \log(\Theta_2) \right)} \right)^2$$

in case of the 2x2 cross-over design, where z_p is the p quantile of the standard normal distribution. 1-β is the power. If $\Theta_1 = 1$ then $z_{(1-\text{ß})}$ has to be replaced by $z_{(1-\text{ß}/2)}$. b_k is the so-called design constant, which is =2 in case of a 2x2 cross-over.

Implementation details

Owen's Q function is implemented via the `integrate()` function of the R package *stats* wich performes numerical integration via an adaptive algorithm.
The function to integrate over is hidden in the internal function

    .Q.integrand(x,nu,t,delta)

To avoid numerical overflow in the factor bevor the definite integral it is calculated logarithmically within that function as

    lnQconst <- -((nu/2.0)-1.0)*log(2.0) - lgamma(nu/2.)

`lgamma(x)` is the $\log(\Gamma(x))$ function from the R package *stats*.

The factor $\sqrt{2\pi}$ vanishes if the density function $\varphi(x)$ of the standard normal distribution in equation (III) is replaced by `exp(-0.5*x^2)`.

Since for really large values of nu and the upper integration limit R the integrand is a function which is is zero over nearly all its range, the `integrate()` function breaks down (see

`help(integrate)`) and `OwensQ()` returns erroneously 0. Therefore the upper integration limit is decreased in steps of R/500 until the `.Q.integrand` has a value >0.

For an alternative implementation of the power calculation according to equation (II) see the function `power.equivalence.md()` of the package *MBESS*. Author of that function is Kem F. Philipps.

The **exact power** according to equation (II) is implemented in the hidden internal function `.power.TOST(alpha=0.05, ltheta1, ltheta2, diffm, se, n, df, bk=2)`. This function is used by the high level functions `power.TOST()` or `sampleN.TOST()` if you set `exact=TRUE` (the default).

The **approximate power** according to the non-central t-distribution is implemented in the hidden internal function
`.approx.power.TOST(alpha=0.05, ltheta1, ltheta2, diffm, se, n, df, bk=2)`
This function is used if your set `exact=FALSE` in `power.TOST()` or `sampleN.TOST()`.

The approximation according to equation (V), via "shifted" central t-distributio is implemented in the hidden function
`.approx2.power.TOST(alpha=0.05, ltheta1, ltheta2, diffm, se, n, df, bk=2)`.
This function is not used further ithin the package and contained only for testing purposes.

Both approximations can yield power values <0. In that case the power will be set =0.

To use these internal functions by yourself, you must supply the values diffm=log(Θ_0), theta1= log(Θ_1) and theta2= log(Θ_2) in case of log-transformed evaluation.
n is the sample size, df the degrees of freedom, bk the design constant.
It is highly recommended to use the high level functions `power.TOST()` or `sampleN.TOST()`. They shield you from all the pecularities of the designs and log-transformed or un-transformed evaluation.

If you are interested in more insight in the implementation load down the source code of the package *PowerTOST* and have a look at the code and especially at the comments within it.

References:

Diletti, E., Hauschke, D., and Steinijans, V.W. (1991)
"Sample Size Determination for Bioequivalence Assessment by Means of Confidence Intervals"
International J. of Clinical Pharmacology, Therapy and Toxicology, Vol. 29, 1-8.

Chow S.-C. and J.-P. Liu (2009)
"Design and Analysis of Bioavailability and Bioequivalence Studies"
Third edition
CRC Press, Chapman & Hall, Boca Raton

Hintze J. (2008)
PASS 2008, NCSS
LLC. Kaysville, Utah

Joulious S.A. (2010)
"Sample sizes for Clinical Trials"
CRC Press, Chapman & Hall 2010

Owen, D.B. (1965)
"A Special Case of a Bivariate Non-central *t*-distribution"
Biometrika, 52, 437 -446.

Phillips, K.F. (1990)
"Power of the Two One-Sided Tests Procedure in Bioequivalence,"
J. of Pharmacokin. and Biopharmaceutics, Vol. 18, No. 2, 137 -144.

SAS release 8.2 - The Analyst application
SAS Institute Inc., *Cary NC, USA, 1999*

Schuirman DJ (1987)
"A comparision of the two one-sided tests procedure and power approach for assessing the equivalence of average bioavailability"
J. of Pharmacokin. and Biopharmaceutics 15: 657-680

Westlake W.J. (1981)
"Response to TBL Kirkwood: bioequivalence testing – a need to rethink"
Biometrics 37, 589-594

Westlake W.J. (1988)
"Bioavailability and bioequivalence of pharmaceutical formulation"
In: Peace KE (ed) "Biopharmaceutical statistics for Drug Development"
Marcel Dekker, New York, 329-352